



下载APP



10 | 如何选择一个适合你的编解码器？

2021-12-13 冯建元

《搞定音频技术》

课程介绍 >

**讲述：冯建元**

时长 13:44 大小 12.58M



你好，我是建元。

上节课我们讲了音频编 / 解码器的基本原理。相信你已经对编 / 解码器有了一个整体的了解。其实编 / 解码器中的算法链路还是比较复杂的，自己从头开始设计和调试一个编 / 解码器的研发成本也是非常巨大的。所以我们一般会选择已有的编 / 解码器来使用。

而音频编 / 解码器经过几十年的发展，其实已经有很多成熟的解决方案可以选择。而且不同的场景对实时音频也有不同的要求。比如，音乐场景要求有比较高的采样率；合唱场则需要比较低的延迟等。那具体根据什么标准来选择编解码器呢？这正是我们这节课的[★]点所在。

这节课我们先来看看编 / 解码器选择时需要重点看哪些指标，然后从几个应用场景的角度看看如何挑选编 / 解码器以及如何选择合适的码率来达到我们想要的效果。

音频编 / 解码中常见的指标

音频编 / 解码器需要关注的指标主要包括码率、音质、计算复杂度和延迟这 4 个大项。你可以先结合图 1 来大体了解一下，下面我会详细介绍。

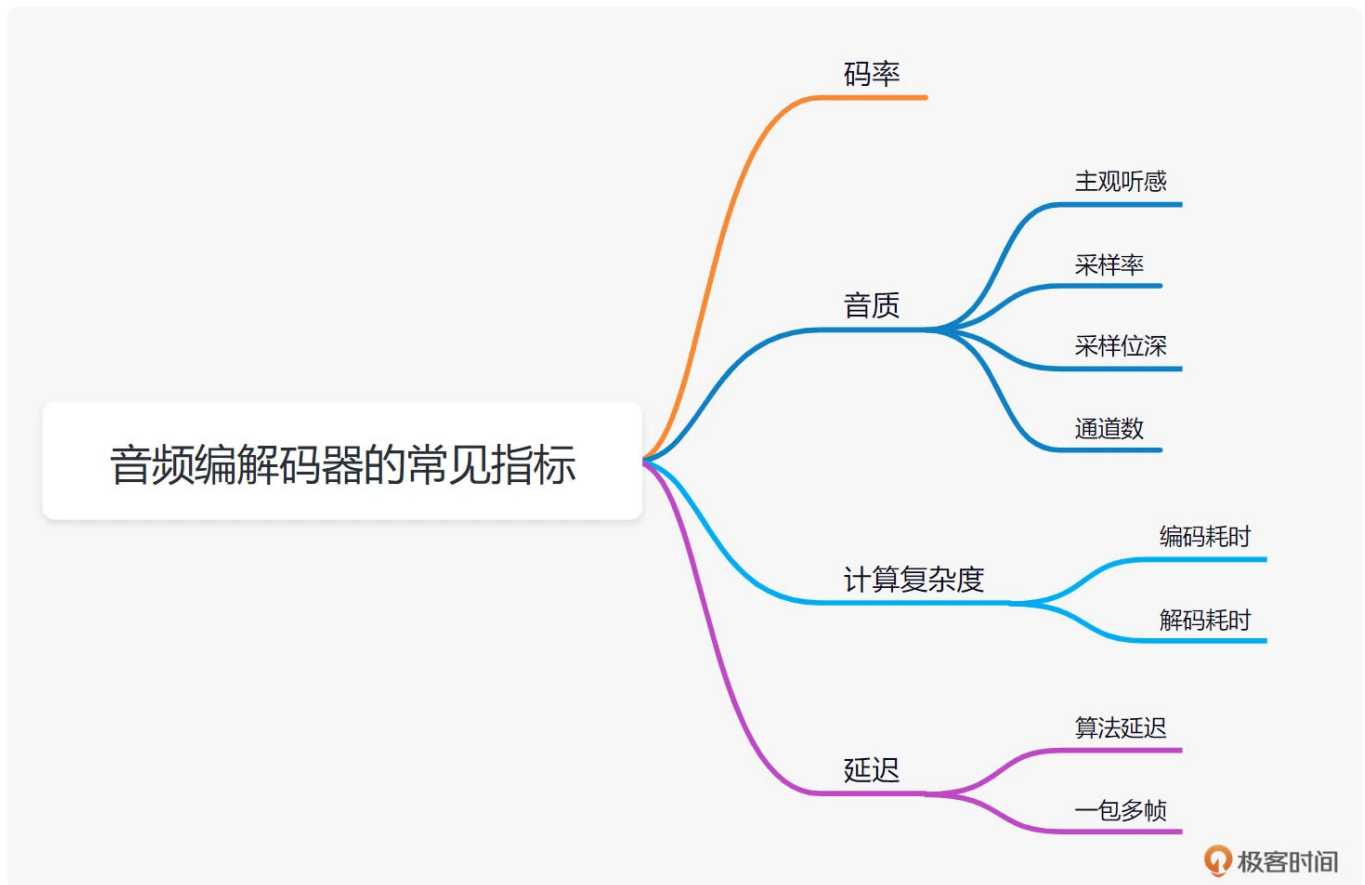


图1 实时音频编/解码的常见指标

码率与音质

音频编 / 解码最直观的目的就是节省传输带宽。所以我们第一个要关注的指标就是码率。现在比较常用的编 / 解码器，比如 OPUS、EVS 等，都是支持不同码率传输的。而不同的码率一般会对应我们要关注的第二项指标音质。你还记得我们之前讲的音频质量评估么，音质我们可以用主观评测试验，也可以用 PESQ、POLQA 等客观评测方法来对主观听感进行打分。当然我们也可以从一些直接的音频指标，比如采样率、采样位深、通道数等来大致衡量音频质量的好坏。

计算复杂度

除了码率和音质，我们下一个要关心的是编 / 解码器的计算复杂度。音频的编码和解码都需要一定的算力支持。你可以回想一下我们之前讲的编 / 解码器的原理，其实大部分常用的编 / 解码器解码的计算力会比编码端的计算力要小很多。一般，我们在多人实时音频互动的时候，其实一个设备需要做自己这一路的编码和多路的解码。所以**在看某个音频编 / 解码是否可用的时候，我们至少需要看看我们的设备是否可以支持实时一路编码和实时多路解码。**

测试方法也比较简单，就是看看在你需要同时互动的最大数量的情况下，**你的终端播放出来的声音是否会出现卡顿或者无声的情况。**

这里从经验上来说，目前移动端的硬件设备，比如说笔记本电脑或者手机，常见的编 / 解码的计算复杂度一般都是可以支持的。但有一些定制化的 IOT 设备或者后台还有其它应用，在同时运行的时候，我们可能需要关注或者测试一下编 / 解码器是否支持你的应用场景。

延迟

在实时音频互动中还有一个需要注意的指标就是延迟。延迟主要包括两个部分：

一个是编 / 解码器算法引入的延迟，比如，编码时依赖未来帧的信息对当前帧进行编码；

另一个是网络发送时组包的延迟，比如，我们把 4 帧作为一个包来发送，那么延迟就会增加 4 帧的时长。而其实我们在音频互动中一般会比较关心一个音频“端到端”的延迟是多少。这里说的端到端，就是从你说的话被麦克风采集传到对端的设备，并从扬声器里播放出来的延迟。**这个端到端的延迟包括了设备采集播放的延迟、音频处理算法引入的延迟、编 / 解码引入的延迟和网络传输的延迟。**

这里我和你分享一些经验：

一般来说“端到端”的延迟如果超过 200ms，人就可以开始感受到音频通话和面对面说话之间的差异。

而如果延迟超过 400ms，那么你可以明显感受到，对面的反映有一种慢半拍的感觉。也就是你说了一句话，对面需要反映一会儿才会给你回应。所以我们在音质、码率等指标

都合适的情况下，如果是需要音频互动的场景，我们一般会选择延迟比较低的编 / 解码器。

好的，说了这么多需要关心的指标，我们可以通过图 2 来看一下不同编 / 解码器的性能差别。

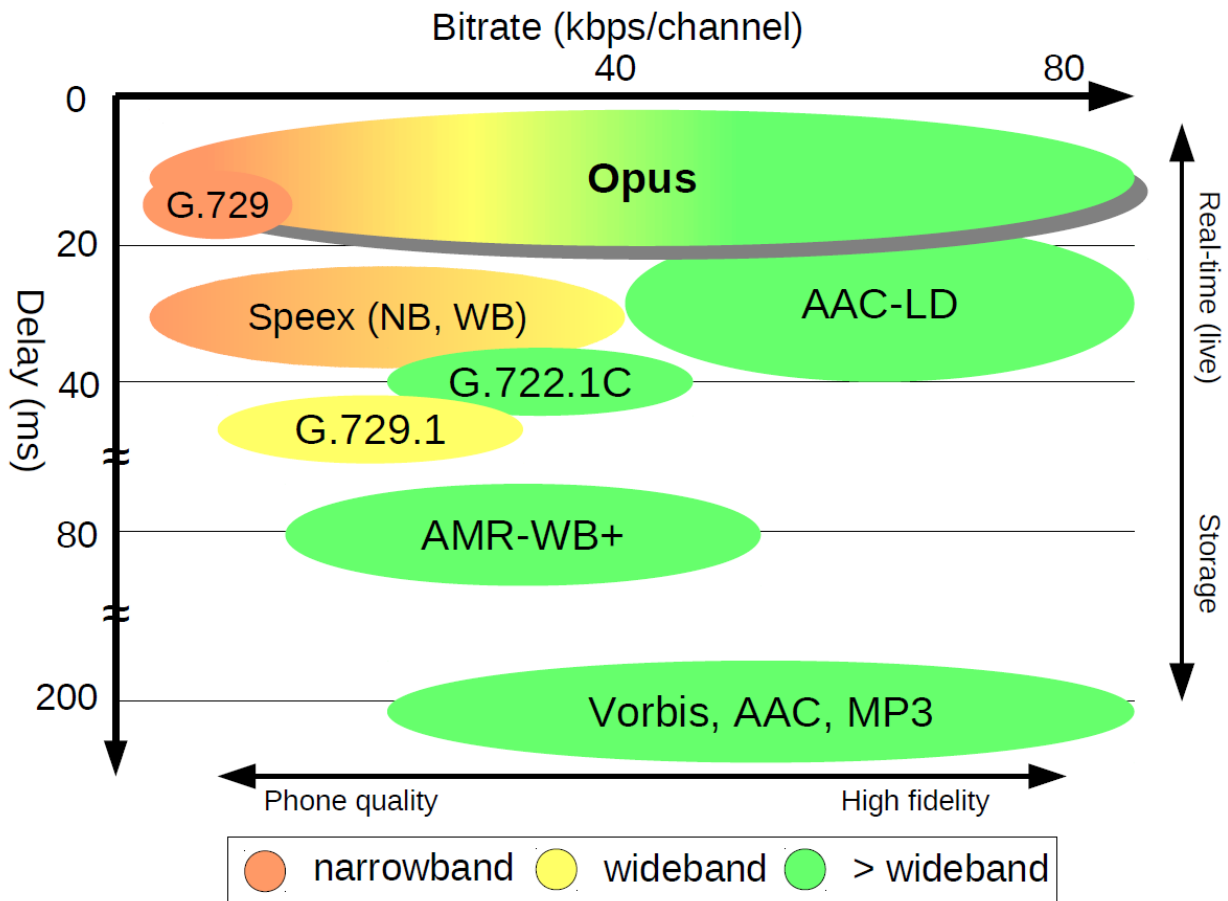


图2 不同编/解码器的性能表现

从这张图中我们可以看到，如果从延迟（Delay）来看，OPUS 和 G.729 是所有编 / 解码器中延迟最小的，可以达到 20ms 左右的延迟。而 OPUS 能提供的码率是 6 ~ 80kbps，覆盖的范围比较大（OPUS 实际可支持的码率可以更高，这里绘图限制在 80kbps 以下的常用码率范围内方便对比）。但要注意这里红黄绿分别代表窄带、宽带和超宽带，而结合上一讲的知识我们知道这和音频的采样率有关。

举个例子，如果你选择了 OPUS 作为编 / 解码器，码率选择了 6kbps，那么你就只能得到一个采样率为 8kHz 的窄带音频。这里需要提醒一下你，这里说的采样率是有效采样率。什么意思呢？也就是说，你可能发现最后解码出来的音频文件是 48kHz，但在频谱图上 4kHz 以上是没有能量的。

通过这张图中 Speex 和 OPUS 的对比，你不难发现 OPUS 的延迟要低于 Speex，同时音质又高于 Speex。可以说 Speex 已经被 OPUS 完全超越了，这也是为什么 Speex 开源项目停止更新的原因。

最后，让我们再来看看图 2 中的 MP3、AAC 这几种编 / 解码器。它们的延迟都达到了 200ms 以上。这并不适合实时音频互动，但它们的存储空间比较少，所以**如果我们需要对音频进行录制或者直播，这种对延迟要求不那么高的情况，可以采用这些音频格式。**

音频编 / 解码器案例分析

音频编 / 解码器的指标还是比较好理解的。但在实际工作中，我们选择编 / 解码器还需要考虑到带宽成本、编 / 解码器是否有开源代码方便集成等问题。所以其实我们做决策的时候，往往需要更为全面地思考问题。这里让我们通过两个真实场景的例子，来看看如何选择适合你的编 / 解码器。

案例 1：在线会议

在线会议，你应该并不陌生。我们来想一下这个场景有什么特点。首先这是一个可能涉及多人的实时互动场景；其次它对音质的要求主要是为了保证语音的通话流畅。

那么转换为音频编 / 解码器的指标来理解，也就是音质首先需要保证语音通话的采样率（比如宽带），其次延迟要小。那么在选择编 / 解码器的时候就可以进行筛选了。

这里举个例子，结合图 2 我们可以看到，如果选择中、低等码率的 OPUS 基本上就可以满足要求了。OPUS 在码率超过 10kbps 之后就可以做到宽带信号的编 / 解码了。

那么具体要选择多少码率呢？这个就需要你对场景业务的理解了。

如果在线会议中可能会有音乐的场景，为了保证音质，你可以把码率调整到比如 64kbps 这种比较高的码率，来保证音乐信号的保真度。

如果你主要是为了保证多人同时互动，而且需要降低带宽成本，那么其实比如 18kbps 的码率，也就可以做到多人语音互动的流畅了。

好了，选择好了码率，现在你可能会好奇，哪些编 / 解码器是开源的呢？

其实为了更好的普及实时音频技术推动标准化的进程。不少行业标准组织、企业联合推出了多款可开源使用的音频编 / 解码器，这里我主要给你介绍比较常用的三个：OPUS、EVS 和 AAC。

其中，OPUS 作为互联网行业最常用的编 / 解码器，是由 Xiph.org 基金会、Skype、Mozilla 基金会联合出品的。OPUS 延迟低、效果也不错，从而在互联网领域有一统江湖之势。并且因为它集成在了 WebRTC 中，所以如果你的应用需要和 Web 端互通音频，那么建议你使用 OPUS。

传统的移动通信标准组织 3GPP 则是 EVS 的幕后推动者。EVS 作为开源的音频编 / 解码器，在音频编 / 解码的各项指标上也可以说是表现优秀。它可以支持 8kHz ~ 48kHz 的音频编码，并支持 5.9kbps 到 128kbps 的码率范围。

而 AAC 编 / 解码器是由 MPEG (Moving Picture Experts Group , 动态图像专家组) 开发的编 / 解码器，其背后是杜比、Sony 等公司的技术集成。是不是听到这些公司的名字你已经有所察觉？没错，AAC 编 / 解码器在音乐编码方面的效果会比较好。AAC 的系列有多个种类，比如 LC-AAC(低复杂性，普通质量) 和 HE-AAC(高效性，高质量) 等。你可以根据计算复杂度、音质等挑选合适的系列。

案例 2：在线 K 歌直播

好的，聊完了会议场景，让我们再来看一个音乐场景：在线 K 歌直播。音乐场景需要高采样率甚至是立体声和多声道。但是不是将码率调高，换一个支持立体声的音频编 / 解码器就解决问题了呢？

越简单的答案，越需要小心。如果你的在线 K 歌直播是在网络覆盖比较差的区域，比如印度，那么你用了一个高码率，比如说 128 kbps 的 EVS 编 / 解码器，来保证音频质量，实际得到的结果可能就是音频卡顿不断，反而影响了效果。这时我们可能会**分地区来使用不同的码率，甚至不同的编 / 解码器。**

除了需要根据地区网络情况来选择编 / 解码器之外，空间感在音乐音质中也是很重要的组成部分。而为了保留空间感，我们最少需要两个声道，也就是我们说的立体声来实现。因此，我们假设网络带宽足够的情况下想要播放双声道的声音，那么可以使用 LC-AAC、HE-AAC 等编 / 解码器的双声道模式，来实现立体声的传输。

那么为了追求极限的音质，我们还可以使用哪些编 / 解码器呢？

为了能比较好地还原音频的空间感，我们往往需要更多的音轨来做到。比如为了还原环绕声的听感，杜比的 AC-3 编 / 解码器可以支持 5.1 声道编 / 解码，也就是说通过 AC-3 可以传输 6 个声道的数据。这样就可以在杜比数码（Dolby Digital）家庭影院系统中播放环绕声了。

一个杜比 5.1 声道的播放系统如图 3 所示。我们可以看到，6 个通道的 AC-3 编码最后分别由 6 个音响进行播放，这样你就可以感受到声音从四周不同的方位传来。

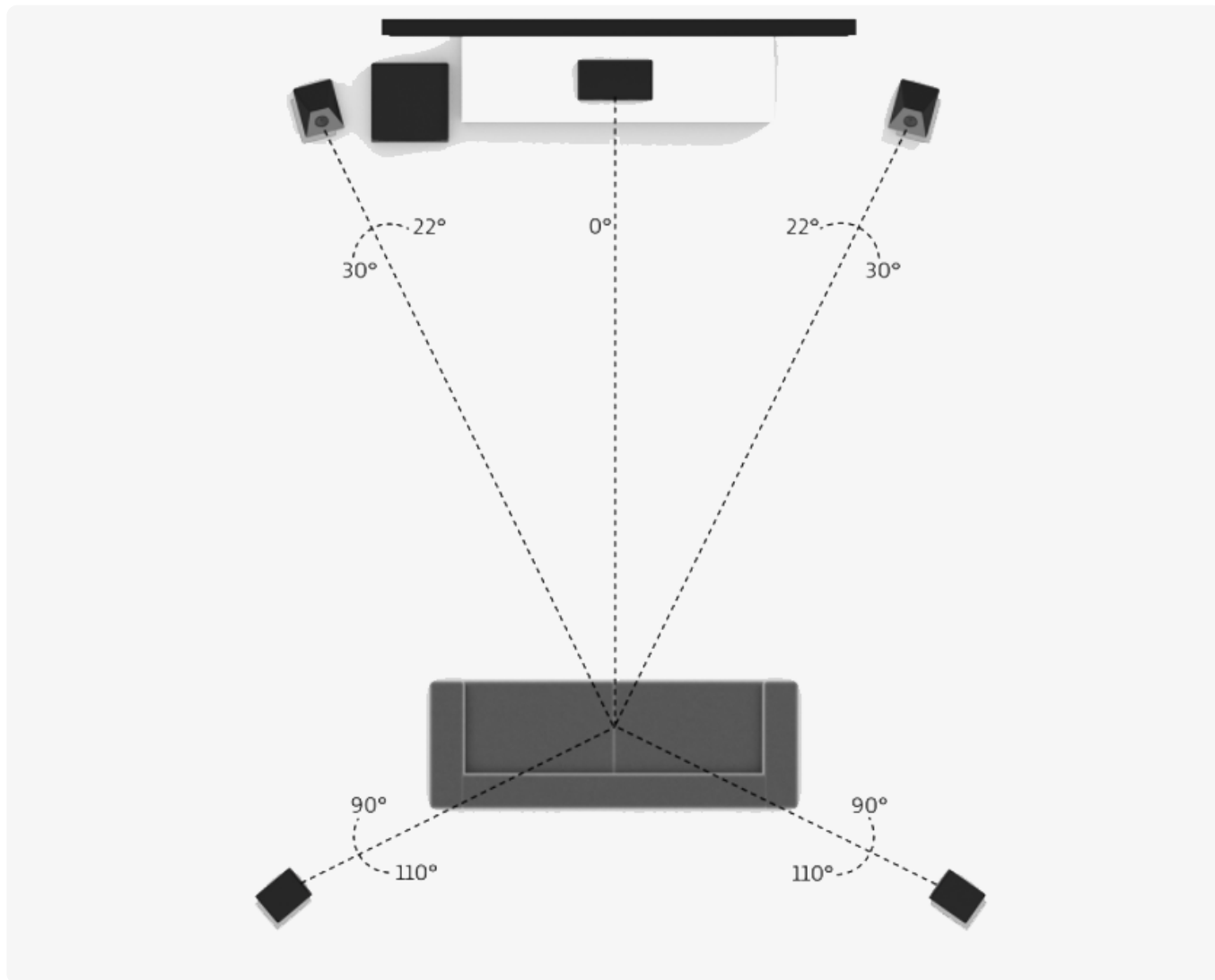


图3 杜比5.1声道播放系统

类似的编 / 解码器还有 DTS 格式。比如你在看电影的时候，如果有标记上有 DTS，就说明这是一个高保真的音频编 / 解码格式。从码率上来说，AC-3 所支持的码率从 32kbps 到 640kbps。而 DTS 的码率比较高，通常为 768 ~ 1536kbps，一般用于 DVD、电影等离线场景。你可以看到由于压缩能力好，AC-3 则在实时互动中更有优势。

小结

好的，今天的课到这里就要结束了。我们来回顾一下这节课的内容。

在你挑选音频编 / 解码器的时候，常见的需要关心的指标主要是码率、音质、计算复杂度、延迟等。一般来说相同码率的情况下，音频编 / 解码器的音质越好、复杂度越低、延迟越小，编 / 解码就越好。但编 / 解码器的码率和通道数等都是可调节的，且音频编 / 解码器的选择是需要根据不同场景来综合考虑的。

文中我给出了两个具有代表性的案例：

会议场景，主要追求的是互动的流畅，所以我们会选择延迟较小的、音质一般的音频编 / 解码器，比如 OPUS。

音乐场景，除了高码率保证音质外，还需要多声道来保证音频的空间感，我们可以使用双声道、甚至多声道的编 / 解码器，例如 AAC 来实现。

如果你对音频中如何还原空间感还有疑问，不要着急，我会在后面的课程为你详细介绍。

其实，最近还有很多基于 AI 的音频编 / 解码器相继发布，可以说基于 AI 技术的编 / 解码器在如何使得码率更小、音质更高上是一个不错的研究方向。

例如，Google 发布的 Lyra，Sound stream，以及微软发布的 Satin。另外，我在声网也研发了国内第一个基于 AI 的语音编 / 解码器  Silver，它可以用 3.6kbps 实现 32kHz 采样率的语音的编 / 解码。你可以看到基于 AI 的编 / 解码器可以把码率降到一个很低的水平，同时又保持较高的音质。有兴趣你可以通过链接自行了解一下。

思考题

最后，这里有一道思考题留给你：很多社交场景中语音和音乐是同时存在的，那么我们有没有什么办法能同时满足沟通的流畅和音乐的音质保真呢？

欢迎你在留言区和我分享你的思考和疑惑，你也可以把今天所学分享给身边的朋友，邀请他加入探讨，共同进步。下节课再见。

分享给需要的人，Ta 订阅后你可得 **20 元现金奖励**



生成海报并分享

👍 赞 1

🔗 提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 09 | 音频编解码器是如何工作的？

下一篇 11 | 网络差怎么办？音频网络传输与抗弱网策略

精选留言

💬 写留言

由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。